Formats of HFC Data Files

Contents

Introduction

At present, the HFC uses only one type of raw data – age-specific fertility rates (ASFR) for all birth orders combined and, when available, by birth order. On the basis of the ASFRs, cumulative fertility rates, total fertility rates, and mean ages at birth are calculated. Details on data adjustment and calculation procedures are proved in the *HFC Methods Protocol*.

In order to better suit the research interests and needs of data users, the HFC presents fertility data in different configurations: in their original format (i.e., PDF, Excel or other type of file in which data were obtained) as well as in a standardized HFC format; with an original age scale and with a standardized single-year age scale; with an original and standardized birth order range; by single source and in pooled (multiple-source) data files; for each country separately and for the entire HFC. Data for all birth orders combined and by birth order are presented in separate data files. Female and male fertility data are also provided separately on the web.

The HFC data files (except the raw data files) are formatted, comma-delimited text files that are easily readable in Excel and statistical packages. A summary of file names can be found in Appendix 1. The first row in each data file contains the header. Missing values are coded as dots ('.'). The age-specific fertility rates are provided to five decimal places, the cumulative fertility rates and the total fertility rates are shown to three decimal places, and the mean ages at birth are provided with the accuracy of two digits after the decimal point.

To facilitate rapid downloads, large pooled HFC data files are compressed in a zip archive.

1. Structure of pooled and single-source data files

The following pooled, or multiple-source, data files that have been created separately for women and men are available for each country. They provide the following data:

1. Adjusted age-specific fertility rates (ASFR) and calculated cumulative period fertility rates (CPFR) with a standardized age scales ≤14, 15, 16, …, 49, 50+ (for women) and ≤15, 16, … , 59, 59+ (for men) and, when available, a standardized birth order range 1, 2, 3, 3+, 4, 4+, 5+. The uniform age scale is obtained and, if applicable, the birth orders are aggregated using a uniform set of HFC methods (for methodological details, see the *HFC Methods Protocol*).

2. Total fertility rates (TFR) and mean ages at birth (MAB) calculated on the basis of the adjusted ASFRs (see the *HFC Methods Protocol*).

These pooled data files in zip archives are made available also for the entire HFC but split for women and men.

Additional pooled data file (consisting separately of female or male fertility data) is available for the entire HFC and includes:

3. Original age-specific fertility rates with an original age scale and birth order range (up to 5+) organized in the standard HFC format.

The single-source data files are also available for men and women and present age-specific fertility rates (in the standard HFC format) by each data source separately. The age scale and the birth order range in these files are preserved as in the original data source. As in the case of pooled data, data for all the birth orders combined and by birth order are placed into separate files.

For the names of all the HFC data files, please see Appendix 1.

1.1 Files for age-specific fertility rates and cumulative fertility rates

The headings of pooled data files:

Heading of pooled data files on the adjusted ASFRs and the CPFRs for all the birth orders combined:

Country, Region, Urban, Origin, Year1, Year2, Age, AgeInt, AgeDef, Vitality, ASFR, CPFR, Collection, SourceType, RefCode, Note, Split

Heading of pooled data files on the adjusted ASFRs and the CPFRs by birth order:

Country, Region, Urban, Origin, Year1, Year2, Age, AgeInt, AgeDef, Vitality, ASFR, ASFR1, ASFR2, ASFR3, ASFR3P, ASFR4, ASFR4P, ASFR5P, CPFR, CPFR1, CPFR2, CPFR3, CPFR3P, CPFR4, CPFR4P, CPFR5P, Collection, SourceType, RefCode, Note, Split

Heading of pooled data files on the original ASFRs for all the birth orders combined (for the entire HFC only):

Country, Region, Urban, Origin, Year1, Year2, Age, AgeInt, AgeDef, Vitality, ASFR, Collection, SourceType, RefCode, Note

Heading of the pooled and single source data files on the original ASFRs by birth order (for the entire HFC only):

Country, Region, Urban, Origin, Year1, Year2, Age, AgeInt, AgeDef, Vitality, ASFR, ASFR1, ASFR2, ASFR3, ASFR3P, ASFR4, ASFR4P, ASFR5P, Collection, SourceType, RefCode, Note

The headings of single-source data files:

Heading of single-source data files on the ASFRs for all the birth orders combined:

Country, Region, Urban, Origin, Year1, Year2, Age, AgeInt, AgeDef, Vitality, ASFR, Collection, SourceType, RefCode, Note

Heading of single-source data files on the ASFRs by birth order:

Country, Region, Urban, Origin, Year1, Year2, Age, AgeInt, AgeDef, Vitality, ASFR, ASFR1, ASFR2, ASFR3, ASFR4,…[1], Collection, SourceType, RefCode, Note

The description of the fields:

1. *Country* (three-character alphabetic): ISO 3166-1 alpha-3 (or equivalent) codes for countries or areas. A description of all the codes used in the HFC is provided in the "Data" section on the HFC website.

2. *Region* (alphanumeric): ISO 3166-2 (or equivalent) region codes for identifying the principal subdivisions (e.g., provinces or states) of countries. When data pertain to the whole country, this field contains a dot ('.'). A description of all codes used in the HFC is provided in the "Data" section on the HFC website.

---

[1] The birth order range in this file is identical to that in the original data source. Therefore the number of fields featuring the ASFRs varies from one single-source file to another.

3. *Urban* (numeric; '.' for Total, '1' for Urban, '2' for Rural): urban/rural coverage of data (when available).

4. *Origin* (numeric): code for ethnicity, religion or race. A dot ('.') denotes the total population.

5. *Year1* (four-digit numeric): the first year of the period to which the data pertain.

6. *Year2* (four-digit numeric): the last year of the period to which the data pertain. If data belong to a single calendar year, the fields "Year1" and "Year2" are filled with the same values.

7. *Age* (numeric): age of the mother/father. For the age groups other than the lower open age interval (*AgeInt* value '-99'), the value of *Age* is the lower age limit.

8. *AgeInt* (numeric; '-99' and '99' for the lower and the upper open age intervals, respectively): the length of the age interval. It is noteworthy that sometimes the data sources do not report whether the first and the last intervals are opened or closed. In the HFC, the interval is shown as open only if it is clearly stated in the data source, otherwise it is closed.

9. *AgeDef* (three- or four-character; 'ACY' for the age in completed years or age at last birthday, 'ARDY' for the age reached during the year, '.' for cases when the data shape is not clearly defined in the source): code for the definition of age used in the original data source.

10. *Vitality* (one-digit numeric; '1' for live births, '2' for total births, '.' for unknown): vitality of births used in the calculation of ASFRs. The vitality of births is indicated only if it is clearly stated in the original data source, otherwise '.' is entered in this field.

11. *ASFR* (numeric; '.' for missing values): age-specific fertility rate for all the birth orders combined.

12. *ASFR1* (numeric; '.' for missing values): age-specific fertility rates for birth order 1.

13. *ASFR2* (numeric; '.' for missing values): age-specific fertility rates for birth order 2.

14. *ASFR3* (numeric; '.' for missing values): age-specific fertility rates for birth order 3.

15. *ASFR3P* (numeric; '.' for missing values): age-specific fertility rates for birth orders 3+. Please note that in cases when data on higher birth orders than the birth order 3 are available in the original data source, the adjusted age-specific fertility rates for birth orders 3+ are calculated by the HFC team.

16. *ASFR4* (numeric; '.' for missing values): age-specific fertility rates for birth order 4.

17. *ASFR4P* (numeric; '.' for missing values): age-specific fertility rates for birth orders 4+. Please note that in cases when data on higher birth orders than the birth order 4 are

available in the original data source, the adjusted age-specific fertility rates for birth orders 4+ are calculated by the HFC team.

18. *ASFR5P* (numeric; '.' for missing values): age-specific fertility rates for birth orders 5+. Please note that in cases when data on higher birth orders than the birth order 5 are available in the original data source, the age-specific fertility rates for birth orders 5+ are calculated by the HFC team.

19. *CPFR* (numeric; '.' for missing values): cumulative fertility rates for all the birth orders combined.

20. *CPFR1* (numeric; '.' for missing values): cumulative fertility rates for birth order 1.

21. *CPFR2* (numeric; '.' for missing values): cumulative fertility rates for birth order 2.

22. *CPFR3* (numeric; '.' for missing values): cumulative fertility rates for birth order 3.

23. *CPFR3P* (numeric; '.' for missing values): cumulative fertility rates for birth order 3+ (see also the explanation for the field *ASFR3P*).

24. *CPFR4* (numeric; '.' for missing values): cumulative fertility rates for birth order 4.

25. *CPFR4P* (numeric; '.' for missing values): cumulative fertility rates for birth order 4+ (see also the explanation for the field *ASFR4P*).

26. *CPFR5P* (numeric; '.' for missing values): cumulative fertility rates for birth order 5+ (see also the explanation for the field *ASFR5P*).

27. *Collection* (alphabetic): code indicating the source of data, i.e. distinguishing data that come from statistical offices and statistical publications, other large data collections, estimates acquired from individual researchers, etc. Meaningful abbreviations are used for these codes (e.g., ODE stands for *L'Observatoire Démographique Européen,* RE denotes data provided by individual researchers, etc.). A description of all such source codes is available in the "Data" section on the HFC website.

28. *SourceType* (alphabetic): code describing the origin of estimates for every source, e.g. fertility data estimated from vital registration system, from census, individual research estimates, etc. A description of all such source codes is provided in the "Data" section on the HFC website.

29. *RefCode* (alphanumeric): reference code. It has format *XXX_##,* in which the first three characters represent the country or area code and the second part is the number of the respective bibliographic reference in the list of references. For example, AUT_01, AUT_02, etc. A list of references is provided both for each country separately (on a country page) and for the entire HFC (in the "Data" section on the HFC website). A combined list of references is created for female and male fertility

data with the consecutive numbering within the country. For the TFRs and the MABs, references to the sources of the ASFRs used to calculate them are provided.

30. *Note* (numeric or '.'): code that links to specific notes in the Notes file. There is a notes file for each country separately (on a country page) and for the entire HFC (in section "Data" on the HFC website). However, the numeration of notes is consecutive within the entire HFC, including female and male fertility data

31. *Split* (one-digit numeric; '1' for Calibrated spline estimator, '2' for Hermite spline, and '0' for no adjustments): code that indicates whether any age splitting procedure was applied to the original data on ASFRs.

1.2 Files for total fertility rates and mean ages at birth

Heading of pooled data files on the TFRs and the MABs for all the birth orders combined:

Country, Region, Urban, Origin, Year1, Year2, AgeDef, Vitality, TFR, MAB, Collection, SourceType, RefCode, Note, Split

Heading of pooled data files on the TFRs and the MABs by birth order:

Country, Region, Urban, Origin, Year1, Year2, AgeDef, Vitality, TFR, TFR1, TFR2, TFR3, TFR3P, TFR4, TFR4P, TFR5P, MAB, MAB1, MAB2, MAB3, MAB3P, MAB4, MAB4P, MAB5P, Collection, SourceType, RefCode, Note, Split

The description of the fields:
1. *Country*: the same as for the files on the ASFRs and the CPFRs.
2. *Region*: the same as for the files on the ASFRs and the CPFRs.
3. *Urban*: the same as for the files on the ASFRs and the CPFRs.
4. *Origin*: the same as for the files on the ASFRs and the CPFRs.
5. *Year1*: the same as for the files on the ASFRs and the CPFRs.
6. *Year2*: the same as for the files on the ASFRs and the CPFRs.
7. *AgeDef*: the same as for the files on the ASFRs and the CPFRs.
8. *Vitality*: the same as for the files on the ASFRs and the CPFRs.
9. *TFR*: total fertility rates for all the birth orders combined.
10. *TFR1*: total fertility rates for birth order 1.
11. *TFR2*: total fertility rates for birth order 2.
12. *TFR3*: total fertility rates for birth order 3.

13. *TFR3P*: total fertility rates for birth order 3+ (for additional details, see the explanation for the field *ASFR3P* above).

14. *TFR4*: total fertility rates for birth order 4.

15. *TFR4P*: total fertility rates for birth order 4+ (for additional details, see the explanation for the field *ASFR4P* above).

16. *TFR5P*: total fertility rates for birth order 5+ (for additional details, see the explanation for the field *ASFR5P* above).

17. *MAB*: mean ages at birth for all the birth orders combined.

18. *MAB1*: mean ages at birth for birth order 1.

19. *MAB2*: mean ages at birth for birth order 2.

20. *MAB3*: mean ages at birth for birth order 3.

21. *MAB3P*: mean ages at birth for birth order 3+ (for additional details, see the explanation for the field *ASFR3P* above).

22. *MAB4*: mean ages at birth for birth order 4.

23. *MAB4P*: mean ages at birth for birth order 4+ (for additional details, see the explanation for the field *ASFR4P* above).

24. *MAB5P*: mean ages at birth for birth order 5+ (for additional details, see the explanation for the field *ASFR5P* above).

25. *Collection*: the same as for the files on the ASFRs and the CPFRs.

26. *SourceType*: the same as for the files on the ASFRs and the CPFRs.

27. *RefCode*: the same as for the files on the ASFRs and the CPFRs.

28. *Note*: the same as for the files on the ASFRs and the CPFRs.

29. *Split*: the same as for the files on the ASFRs and the CPFRs.

2. Raw data files

These files provide the original data in the format obtained from the data provider/source. For example, if data came from a publication, scanned copies of the respective pages are provided in a PDF format. If data were obtained in an Excel file, the original Excel file is made available for HFC data users. If there is more than one file (or the file is large), they are compressed into a zip archive. Some raw data are accompanied by a methodological description or documentation; this file is then added to the zip archive.

The name of a file for raw data consists of a country code and respective reference code: XXX_## in the case of data for women. For men it contains *m_* in the beginning of the name to distinguish the male fertility data (m_XXX_##).

Acknowledgements

APPENDIX 1: Data available in the HFC and names of files

The three characters in the file name denoted as 'XXX' in this document stand for a standard (uppercase) ISO code, representing a country or area name (AUT for Austria, CAN for Canada, etc.). For countries, the ISO 3166-1 alpha-3 codes are used (see, for example, Wikipedia: http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3). For sub-populations or areas not corresponding to an ISO country, the ISO 3166-2 region codes are used (where available) as an extension of the country code (see http://en.wikipedia.org/wiki/ISO_3166-2).

Table A1: Types and names of data files on a **country page**

All birth orders combined	By birth order
Female fertility data	
Pooled (multiple-source) data files	*Pooled (multiple-source) data files*

1. ASFRs and CPFRs, standardized age scale
  Name: **XXX_ASFRstand_TOT.txt**
	1. ASFRs and CPFRs by birth order, standardized age scale
	  Name: **XXX_ASFRstand_BO.txt**

2. Total fertility rate and mean age at birth
  Name: **XXX_TFRMAB_TOT.txt**
	2. Total fertility rate and mean age at birth by birth order
	  Name: **XXX_TFRMAB_BO.txt**

Single-source data files	*Single-source data files*

3. ASFRs, original age scale; split by the source
  Name: **XXX_##_TOT.txt**
	3. ASFRs by birth order, original age scale; split by the source
	  Name: **XXX_##_BO.txt**

4. Raw data files
  Name:
XXX_##.csv, XXX_##.xls, XXX_##.txt, XXX_##.pdf, or XXX_##.zip
	4. Raw data files
	  Name:
	XXX_##.csv, XXX_##.xls, XXX_##.txt, XXX_##.pdf, or XXX_##.zip

Male fertility data	
Pooled (multiple-source) data files	*Pooled (multiple-source) data files*

1. ASFRs and CPFRs, standardized age scale
  Name: m_**XXX_ASFRstand_TOT.txt**
	1. ASFRs and CPFRs by birth order, standardized age scale
	  Name: m_**XXX_ASFRstand_BO.txt**

2. Total fertility rate and mean age at birth
  Name: m_**XXX_TFRMAB_TOT.txt**
	2. Total fertility rate and mean age at birth by birth order
	  Name: m_**XXX_TFRMAB_BO.txt**

Single-source data files	*Single-source data files*

3. ASFRs, original age scale; split by the source
  Name: m_**XXX_##_TOT.txt**
	3. ASFRs by birth order, original age scale; split by the source
	  Name: m_**XXX_##_BO.txt**

4. Raw data files
  Name:
m_XXX_##.csv, m_XXX_##.xls, m_XXX_##.txt, m_XXX_##.pdf, or m_XXX_##.zip
	4. Raw data files
	  Name:
	m_XXX_##.csv, m_XXX_##.xls, m_XXX_##.txt, m_XXX_##.pdf, or m_XXX_##.zip

Table A2: Names of files containing all the HFC data on the main HFC page

All the birth orders combined	By birth order
Female fertility data	
Pooled (multiple-source) data files	*Pooled (multiple-source) data files*
1. ASFRs and CPFRs, standardized age scale 　Name: **HFC_ASFRstand_TOT.txt** in a zip archive	1. ASFRs and CPFRs by birth order, standardized age scale 　Name: **HFC_ASFRstand_BO.txt** in a zip archive
2. ASFRs, original age scale 　Name: **HFC_ASFR_TOT.txt** in a zip archive	2. ASFRs, original age scale 　Name: **HFC_ASFR_TOT.txt** in a zip archive
3. Total fertility rate and mean age at birth 　Name: **HFC_TFRMAB_TOT.txt** in a zip archive	3. Total fertility rate and mean age at birth by birth order 　Name: **HFC_TFRMAB_BO.txt** in a zip archive
Male fertility data	
Pooled (multiple-source) data files	*Pooled (multiple-source) data files*
1. ASFRs and CPFRs, standardized age scale 　Name: **m_HFC_ASFRstand_TOT.txt** in a zip archive	1. ASFRs and CPFRs by birth order, standardized age scale 　Name: **m_HFC_ASFRstand_BO.txt** in a zip archive
2. ASFRs, original age scale 　Name: **m_HFC_ASFR_TOT.txt** in a zip archive	2. ASFRs, original age scale 　Name: **m_HFC_ASFR_TOT.txt** in a zip archive
3. Total fertility rate and mean age at birth 　Name: **m_HFC_TFRMAB_TOT**.txt in a zip archive	3. Total fertility rate and mean age at birth by birth order 　Name: **m_HFC_TFRMAB_BO.txt** in a zip archive